# Additional Aspects of BLUP

Peter von Rohr

2022-11-25

* Prediction of breeding values using linear mixed effects models.
* Examples: Sire model, animal model
* Solutions for fixed effects estimates and for predicted breeding values can be obtained using Mixed Model equations. In MME, A^{-1} required ==> efficient computations shown

# Aspects

- Accurracy
  - Results from MME are estimates of fixed effects and predictions of breeding values
  - Need statement about quality of estimates and predictions
- Confidence Intervals
- Decomposition of Predicted Breeding values

True value for u_i is with probability (1-alpha) between l and r

0.99
0.95

With estimates and predictions, there are always errors associated ==> quantify error

# Accurracy

fixed effects $\beta - \hat{\beta}$

breeding values $u - \hat{u}$

- One property of BLUP was that variance of prediction error is minimal
- How can we measure the variance of the prediction error
- Fixed effects

fixed $\Rightarrow$ $var(\beta) = 0$

$cov(\beta, \hat{\beta}) = 0$

$$var(\beta - \hat{\beta}) = \underbrace{var(\hat{\beta})}_{\text{e.g.: } (X'X)^{-1} \sigma_e^2}$$

- Random effects

$\rightarrow$ $var(u) \neq 0$

$$var(u - \hat{u}) = var(u) - 2 * cov(u, \hat{u}) + var(\hat{u}) = var(u) - var(\hat{u}) = \overbrace{PEV(\hat{u})}^{min}$$

because with BLUP: $cov(u, \hat{u}) = var(\hat{u})$

Aim: predicted breeding values should be as close as possible to the true breeding values ==> PEV ==> 0

# PEV

In practise: Coefficient matrix of MME cannot be inverted, hence PEV are always approximation

- PEV depends on inverse of coefficient matrix of MME

$$\left[\begin{array}{cc} X^TR^{-1}X & X^TR^{-1}Z \\ Z^TR^{-1}X & Z^TR^{-1}Z + G^{-1} \end{array}\right]^{-1} = \left[\begin{array}{cc} C^{11} & C^{12} \\ C^{21} & \boxed{C^{22}} \end{array}\right]$$

- For predicted breeding values $\hat{u}$

$$PEV(\hat{u}) = var(u) - var(\hat{u}) = C^{22}$$

# Single Animal *i*

$$\begin{array}{c|c} \begin{bmatrix} C^{11} & C^{12} \\ C^{21} & \boxed{C^{22}} \end{bmatrix} & \begin{matrix} 1 \\ 2 \\ \vdots \\ i \end{matrix} \end{array}$$

$$var(u_i) - var(\hat{u}_i) = PEV(\hat{u}_i) = \underline{(C)^{22}_{ii}}$$

i-th diagonal element

$PEV(\hat{u})$

where $(C)^{22}_{ii}$ is the *i*-th diagonal of $C^{22}$

- Accuracy measured by correlation

BLUP : $cov(u_i, \hat{u}_i) = var(\hat{u}_i)$

$$\underbrace{r_{u_i,\hat{u}_i} = \frac{cov(u_i,\hat{u}_i)}{\sqrt{var(u_i) * var(\hat{u}_i)}}}_{\text{Def } r()} = \underbrace{\sqrt{\frac{var(\hat{u}_i)}{var(u_i)}}}_{r^2 = \frac{var(\hat{u}_i)}{var(u_i)}}$$

- Combining

$$PEV(\hat{u}_i) = (C)^{22}_{ii} = var(u_i) - \overset{r^2 \cdot var(u_i)}{var(\hat{u}_i)} = var(u_i) - r^2_{u_i,\hat{u}_i} var(u_i)$$

$$(1 - r^2)\, var(u_i)$$

# ~~Accuracy~~ $B_i$

Reliability

Bestimmtheitsmass

small: PEV → 0
⇒ B → 1

$$B_i = r^2_{u_i,\hat{u}_i} = \frac{var(u_i) - (C)^{22}_{ii}}{var(u_i)} = 1 - \frac{\boxed{PEV(\hat{u}_i)}}{var(u_i)} = 1 - \frac{(C)^{22}_{ii}}{var(u_i)}$$

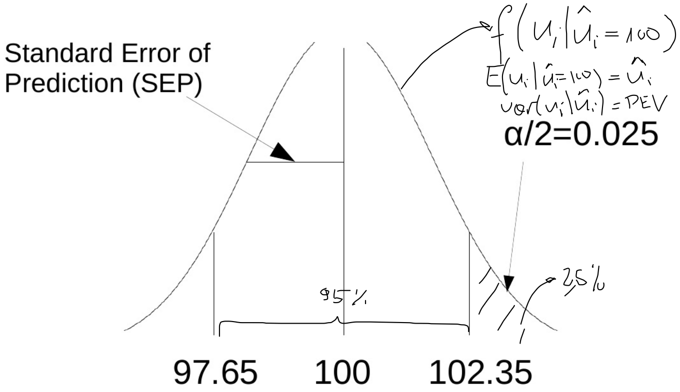
- $B_i$ is large for small $PEV(\hat{u}_i)$
- In the limit $B_i \rightarrow 1$ for $PEV(\hat{u}_i) \rightarrow 0$
- For $PEV(\hat{u}_i) \rightarrow 0$ we must have $var(\hat{u}_i) \rightarrow var(u_i)$
- Therefore, the closer $var(\hat{u}_i)$ is to $var(u_i)$, the more accurate the predicted breeding value

# Confidence Intervals of $\hat{u}_i$

- Predicted breeding value ($\hat{u}_i$) is a function of the data ($y$)
- Hence $\hat{u}_i$ is a random variable with a distribution

# Distribution

SEP: Standard error of prediction

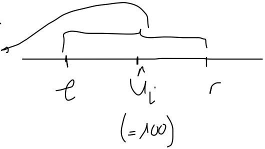


$$SEP(\hat{u}_i) = \sqrt{PEV(\hat{u}_i)} = \sqrt{(1 - \underbrace{r^2_{u_i,\hat{u}_i}}) * \underbrace{var(u_i)}_{\Lambda 2}}$$

# Widths Of Confidence Intervals

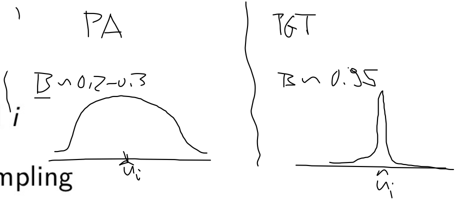
Table 1: Widths of Confidence Intervals for Given Accuracies

| B ~~Accurracy~~ | Interval Width |
|---:|---:|
| → 0.40 | 36.44 |
| 0.50 | 33.26 |
| 0.60 | 29.75 |
| 0.70 | 25.76 |
| 0.80 | 21.04 |
| 0.90 | 14.88 |
| 0.95 | 10.52 → 95 - 105 |
| 0.99 | 4.70 |



with $\hat{u}_i = 100$, $var(u_i) = 144$ and $\alpha = 0.05$ $\Longrightarrow$ 95% CI

# Selection Response



Correlation $r_{u_i,\hat{u}_i}$ for a single animal $i$

PEV

- across conceptual repeated sampling
- change of a predicted breeding value for animal $i$ with increasing information
- related to standard error of prediction (SEP) → measure of risk of using $i$ as parent

---

Accuracies also important for selection response $R = i \cdot r \cdot \sigma_g$

- correlation between true and predicted breeding values in selection candidates
- characteristic of population not of single animal
- estimation with cross-validation

population accuracy

→ use correlation between predicted breeding values with whole and partial data

Recap 2022-12-02:

- Errors for estimates of fixed effects
- Prediction errors, measured Prediction error variance (PEV)
- PEV related to the reliability (Bestimmtheitsmass B%) of predicted Breeding value
- Reliability (B) corresponds to squarred correlation (r) between true and predicted breeding value

$$B_i = r(u_i, \hat{u}_i)^2$$

- B can be computed based on PEV and the PEV is obtained from the inverse of the coefficient Matrix of the MME

- Square root of PEV corresponds to the Standard Error of Prediction (SEP).
- SEP can be used to compute confidence intervalues of the predicted breeding values for a desired error level

Today: Decomposition of the predicted breeding value, to show the different influences and the sources of information that are used in the predicted breeding values.

# Decomposition of Predicted Breeding Value

BLUP uses all info

$$\underbrace{\begin{bmatrix} X^TR^{-1}X & X^TR^{-1}Z \\ Z^TR^{-1}X & Z^TR^{-1}Z + G^{-1} \end{bmatrix}}_{M} \underbrace{\begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix}}_{s} = \underbrace{\begin{bmatrix} X^TR^{-1}y \\ Z^TR^{-1}y \end{bmatrix}}_{r}$$

- Write MME as

Coefficent matrix M

$$M \cdot s = r \longrightarrow$$

right-hand side

with

Herd, sex, lact.

$$s = \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix}$$

Predicted breeding value

$$\hat{u} = \begin{bmatrix} \hat{u}_1 \\ \boxed{\hat{u}_2} \\ \vdots \\ \hat{u}_q \end{bmatrix} \longrightarrow ?$$

- $\hat{\beta}$ has length $p$
- $\hat{u}$ has length $q$

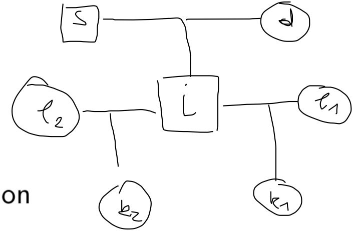
# Simplified Model

Common mean as the only fixed effect, replacement for more complex fixed effect structures

$$y_i = \mu + u_i + e_i$$

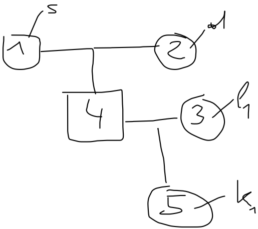
| | | |
|---|---|---|
| where | $y_i$ | Observation for animal $i$ |
| | $u_i$ | breeding value of animal $i$ with a variance of $(1 + F_i)\sigma_u^2$ |
| | $e_i$ | random residual effect with variance $\sigma_e^2$ |
| | $\mu$ | single fixed effect |

# Data



- all animals have an observation
- animal $i$ has
  - parents $s$ and $d$
  - $n$ progeny $k_j$ (with $j = 1, \dots, n$)
  - $n$ mates $l_j$ (with $j = 1, \dots, n$).
- progeny $k_j$ has parents $i$ and $l_j$.

# Example



| Animal | Sire | Dam | WWG |
|---|---|---|---|
| 1 | NA | NA | 4.5 |
| 2 | NA | NA | 2.9 |
| 3 | NA | NA | 3.9 |
| 4 | 1 | 2 | 3.5 |
| 5 | 4 | 3 | 5.0 |

$i = 4$

$y$

Variance components $\sigma_e^2 = 40$ and $\sigma_u^2 = 20$.

Question: What type of information is influencing the predicted breeding value of animal 4?

# Model Components

$$y = X\mu + Zu + e$$

$$y = \begin{bmatrix} 4.5 \\ 2.9 \\ 3.9 \\ \boxed{3.5} \\ 5.0 \end{bmatrix} \quad X = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \; Z = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$X^T X = \begin{bmatrix} 5 \end{bmatrix}, \; X^T Z = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

$$Z^T Z = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

# Right-hand Side

$$X^T y = \left[ \sum_{j=1}^{n} y_i \right] = 19.8$$

$$Z^T y = \begin{bmatrix} y1 \\ y2 \\ y3 \\ y4 \\ y5 \end{bmatrix} = \begin{bmatrix} 4.5 \\ 2.9 \\ 3.9 \\ 3.5 \\ 5 \end{bmatrix}$$

# $A^{-1}$

```
## as(<dtTMatrix>, "dtCMatrix") is deprecated since Matrix
```

$$A^{-1} = \begin{bmatrix} 1.5 & 0.5 & 0 & -1 & 0 \\ 0.5 & 1.5 & 0 & -1 & 0 \\ 0 & 0 & 1.5 & 0.5 & -1 \\ -1 & -1 & 0.5 & 2.5 & -1 \\ 0 & 0 & -1 & -1 & 2 \end{bmatrix}$$

# MME

$$\begin{bmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z + A^{-1} * \lambda \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X^T y \\ Z^T y \end{bmatrix}$$

# Insert Data

$X^TX$

$X^TZ$

$\hat{\beta}$

$$\begin{bmatrix} \boxed{5} & 1 & 1 & 1 & 1 & 1 \\ 1 & 4 = 1+3 & 1 & 0 & -2 & 0 \\ 1 & 1 & 4 & 0 & -2 & 0 \\ 1 & 0 & 0 & 4 & 1 & -2 \\ 1 & -2 & -2 & 1 & 6 & -2 \\ 1 & 0 & 0 & -2 & -2 & 5 \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{u}1 \\ \hat{u}2 \\ \hat{u}3 \\ \boxed{\hat{u}4} \\ \hat{u}5 \end{bmatrix} = \begin{bmatrix} 19.8 \\ 4.5 \\ 2.9 \\ 3.9 \\ \boxed{3.5} \\ 5 \end{bmatrix}$$

$\xrightarrow{} X^Ty$

$Z^TX$

$\hat{u}$

$Z^Ty$

$$\underbrace{Z^TZ}_{I} + \underbrace{\lambda}_{} \cdot A^{-1}$$

$$\sigma_e^2/\sigma_u^2 = 40/20 = 2$$

$$\begin{bmatrix} 5 & 1 & 1 & 1 & 1 & 1 \\ 1 & 4 & 1 & 0 & -2 & 0 \\ 1 & 1 & 4 & 0 & -2 & 0 \\ 1 & 0 & 0 & 4 & 1 & -2 \\ 1 & -2 & -2 & 1 & 6 & -2 \\ 1 & 0 & 0 & -2 & -2 & 5 \end{bmatrix} \begin{bmatrix} \mu \\ \hat{u}1 \\ \hat{u}2 \\ \hat{u}3 \\ \hat{u}4 \\ \hat{u}5 \end{bmatrix} = \begin{bmatrix} 19.8 \\ 4.5 \\ 2.9 \\ 3.9 \\ \boxed{3.5} \\ 5 \end{bmatrix}$$

$$1\cdot\hat{\mu} + (-2)\hat{u}_1 + (-2)\hat{u}_2 + 1\cdot\hat{u}_3 + \underbrace{6\cdot\hat{u}_4}_{!} + (-2)\hat{u}_5 = 3.5$$

# Animal 4

- parents 1 and 2
- progeny 5
- mate 3
- inspection of second but last equation in MME where $y_4$ and $\hat{u}_4$ occur
- Remember from construction of $A^{-1}$, the variable $d^{ii}$ can assume the following values

$$d^{ii} = \begin{cases} 2 & \text{both parents known} \\ \frac{4}{3} & \text{one parent known} \\ 1 & \text{both parents unknown} \end{cases}$$

# Extract Equation

$$y_4 = 3.5 = 1 * \hat{\mu} - 2 * \hat{u}_1 - 2 * \hat{u}_2 + 1 * \hat{u}_3 + 6 * \underline{\hat{u}_4} - 2 * \hat{u}_5$$

$$6 \cdot \hat{u}_4 = 3.5 - \hat{\mu} + 2\hat{u}_1 + 2\hat{u}_2 - \hat{u}_3 + 2\hat{u}_5$$

- Solving for $\hat{u}_4$

$$\hat{u}_4 = \frac{1}{6}\Big[\overbrace{\overbrace{\underbrace{y_4}_{} - \underbrace{\hat{\mu}}_{\text{correction for fixed effects}}}^{3.5}}^{y - X\hat{\beta}} + 2 * (\overbrace{\hat{u}_1 + \hat{u}_2}^{\text{predicted breeding values of parents of 4}}) - \underbrace{\hat{u}_3}_{\text{predicted breeding value of mate}} + \overbrace{2\hat{u}_5}^{\text{progeny}}\Big]$$

- $\hat{u}_4$ depends on
  - own performance record $y_4$
  - estimate of fixed effect $\underline{\hat{\mu} \text{ - environment}}$
  - predicted breeding value of parents 1 and 2, mate 3 and progeny 5

Comparison of BLUP Animal model breeding values to other predictions:

* Regression method using own performance

Populationmean

$$\hat{u}_4 = h^2(y_4 - \underline{\mu}) \qquad h^2 = \frac{\sigma_u^2}{\sigma_p^2} = \frac{20}{20+40} = 1/3$$

* Parent offspring

$$\hat{u}_4 = \frac{2n}{n+k}(\underbrace{\bar{y}_4} - \mu)$$

number of progeny of 4

Average of progeny performance of animal 4

$$k = \frac{4-h^2}{h^2}$$

BLUP-AM

$$\hat{u}_4 = \frac{1}{6}\left[\underbrace{y_4 - \hat{\mu}}_{y_4 - X\hat{\beta}} + 2 * (\underline{\hat{u}_1 + \hat{u}_2}) - \hat{u}_3 + 2\underline{\hat{u}_5}\right] \longrightarrow \hat{u}_5 = \frac{1}{\omega}\left[y_5 - X\hat{\beta} \dots\right.$$

performance corrected for individual environment

Summary:

* BLUP-Animal Model based predicted breeding values consider all available information (as examples, see previous slide)
* Only information that are available to breeding organisation and to the person who runs the genetic evaluation can be considered
* Preferential treatment of animals in herds is often difficult to capture.

Animals that are considered to be very valuable by the breeder
are treated different than all other animals in the herd.
Examples for this are potential dams of future bulls

# General Equation

$$\hat{u}_i = \frac{1}{1 + \alpha\delta^{(i)} + \frac{\alpha}{4}\sum_{j=1}^{n}\delta^{(k_j)}} \left[y_i - \hat{\mu} \right.$$
$$\left. + \frac{\alpha}{2}\left\{\delta^{(i)}(\hat{u}_s + \hat{u}_d) + \sum_{j=1}^{n}\delta^{(k_j)}(\hat{u}_{k_j} - \frac{1}{2}\hat{u}_{l_j})\right\}\right]$$

| where | | |
|---|---|---|
| | $\alpha$ | ration between variance components $\sigma_e^2/\sigma_u^2$ |
| | $\delta^{(j)}$ | contribution for animal $j$ to $A^{-1}$ |